Formalizing value-guided argumentation for ethical systems design

Verheij, Bart

Published in:
Artificial Intelligence and Law

DOI:
10.1007/s10506-016-9189-y

Publication date:
2016

Artif Intell Law (2016) 24:387–407
DOI 10.1007/s10506-016-9189-y

Formalizing value-guided argumentation for ethical systems design

Bart Verheij[1]

Published online: 8 November 2016

Abstract The persuasiveness of an argument depends on the values promoted and demoted by the position defended. This idea, inspired by Perelman's work on argumentation, has become a prominent theme in artificial intelligence research on argumentation since the work by Hafner and Berman on teleological reasoning in the law, and was further developed by Bench-Capon in his value-based argumentation frameworks. One theme in the study of value-guided argumentation is the comparison of values. Formal models involving value comparison typically use either qualitative or quantitative primitives. In this paper, techniques connecting qualitative and quantitative primitives recently developed for evidential argumentation are applied to value-guided argumentation. By developing the theoretical understanding of intelligent systems guided by embedded values, the paper is a step towards ethical systems design, much needed in these days of ever more pervasive AI techniques.

Keywords Argumentation · Ethical systems · Teleological reasoning · Values

1 Introduction

Once Artificial Intelligence was science fiction, and the study of ethical AI could be left to creative speculation in novels and films. A good example of a fictional ethical system appears in Verhoeven's 1987 film *Robocop*, where the choices of a cyborg police officer are guided by three 'prime directives':

1. Serve the public trust;
2. Protect the innocent;

✉ Bart Verheij
bart.verheij@rug.nl

[1] Artificial Intelligence, University of Groningen, Groningen, The Netherlands

3. Uphold the law.

These directives—inspired by Asimov's 1942 Three Laws of Robotics—guide Robocop's behavior, but the plot involves several twists where ethical choices based on Robocop's personal values must be made.

Today Artificial Intelligence is a science with real life applications, and the investigation of ethical AI should be done systematically by scientists and engineers. Autonomous systems for driving and warfare must do the right thing in complex, unforeseeable situations. The design of social media asks for a careful balance between what is good for users and for businesses. The invention of virtual currencies and related blockchain-based technology inspires the automation of trust mechanisms in finance and other businesses.

Advanced intelligent techniques operate in problem domains that involve the complex ethical decision-making that people perform routinely everyday. And even though we make many mistakes—often enough with extremely bad consequences—humans outperform all other natural and artificial systems in real-life ethical decision making. Only we can choose our actions while carefully considering the context, taking human values into account, and following normative rules.

The state-of-the-art in artificial systems with ethical impact in use today is what have been called *implicit ethical systems* (Moor 2006): they are limited by design to performing only the right kind of behavior. Think of an ATM that is carefully designed in order to give money only to the person entitled to receiving it. When Silicon Valley speaks of ethical system design today, this typically concerns such implicit ethical systems (see, e.g., the interesting and relevant TEDx talk by Tristan Harris[1] on systems incorporating human values).

In contrast, Moor speaks of *full ethical systems* when they have an embedded explicit ethical model. Such a model allows a system to make autonomous judgments it can justify, in the face of conflicting ethical considerations. There is a slow shift of attention towards full ethical systems, but the technological hurdles are huge and require fundamental research (see also Broersen 2014, who emphasises responsibility in intelligent systems).

We study ethical decision making using an argumentation perspective, focusing on three themes:

Context-dependence *An ethical system's decisions depend on the circumstances.* What counts as a good decision in one situation, may not be good in another, similar situation. Similarities and differences between the circumstances of situations can determine what counts as a good decision. For instance, when driving a car, an abrupt stop can be a good choice to avoid a collision in front of you, but maybe not when someone is close behind you.

Value-dependence *An ethical system's decisions depend on values embedded in the system.* A system's decisions are not determined by the external circumstances alone. There is room for discretionary choices depending on the values embedded in the system. For instance, when driving a car, some base their choices more on speed, others more on safety.

[1] http://www.ted.com/talks/tristan_harris_how_better_tech_could_protect_us_from_distraction.

Rule-dependence *An ethical system's decisions depend on rules embedded in the system.* A system's decisions are typically not made on a case by case basis, but follow rules. For instance, when driving in a residential area, as a rule you reduce your speed. It does not matter much which residential area you are in, not even whether you have been there before.

In this paper, we take inspiration from research on these themes as performed in the field of Artificial Intelligence and Law. In this field—formally started in 1987 when the first International Conference on AI and Law (ICAIL) was held—the study of these three themes goes back to its early days. In particular, in a series of papers, Hafner and Berman developed a modeling perspective on decision making in the law emphasising its context-dependence, value-dependence and rule-dependence (Berman and Hafner 1991, 1993, 1995; Hafner and Berman 2002). As a tribute, the present paper connects to this work, eminent in its balance between legal scholarship and technical creativity. Our technical approach abstracts from the institutional context of the law, and does (for instance) not consider the designated roles of the parties in court and how they are procedurally enabled and bound by the rule of law. By these abstractions, the ethical decision making underlying legal decision making is emphasised, showing the roles of context, values and rules.

Our specific focus concerns the comparison of values and its role in decision making. Values are typically studied using either qualitative or quantitative modeling primitives. For instance, values are modeled as a qualitative logical property that can either be promoted or demoted when a decision is made (as, e.g., in value-based argumentation frameworks by Bench-Capon 2003). Alternatively, values are handled using quantitative numeric properties such as the probability that a consequence follows and the utility of a decision (as, e.g., in expected utility theory; see Briggs 2014).

In recent research on evidential argumentation (Verheij 2014), a model has been developed for the connection between qualitative and quantitative modeling primitives. That model was designed as an answer to lessons learnt in research on the modeling of arguments and scenarios for evidential reasoning in Bayesian networks (Verheij et al. 2016). In that work, we encountered the well-known issues of Bayesian network modeling that Bayesian networks often require many more assumptions about numeric values and dependencies than are reasonably available, and that there is the risk of misinterpreting the graph underlying a Bayesian network (in particular unwarranted causal interpretation; cf. Dawid 1987).

In this paper, the model presented by Verheij (2014) is applied to the comparison of values in ethical decision making, emphasising the role of context-dependence, value-dependence and rule-dependence. In this way, we provide a perspective on ethical decision making as value-guided argumentation.[2]

[2] Verheij (2014) provides a semi-formal presentation of the model. A corresponding formalism was presented at the AI4J 2016 workshop (Verheij 2016c), of which formal properties were presented at JELIA 2016 (Verheij 2016b). The connection to ethical systems design was made at JURIX 2016 (Verheij 2016a). In this paper, the connection to teleological and value-based argumentation in AI and Law is developed.

2 Value-guided argumentation in Artificial Intelligence and Law

As said, in our approach to ethical systems design, we take inspiration from research on value-guided argumentation in Artificial Intelligence and Law. Among the first to recognize the role of values were Hafner and Berman. In their work on case-based argumentation in the law (Berman and Hafner 1991, 1993, 1995; Hafner and Berman 2002), they emphasise the role of social values in the decision making of courts. Such decision making is often purpose-oriented or teleological, in the sense that the purpose of promoting one social value may have to be balanced with the purpose of promoting another, competing value. Hafner and Berman write that legal precedents are 'embedded in a political context, where competing policies and values are balanced by the courts, and where legal doctrines evolve to accommodate new social and economic realities' (Hafner and Berman 2002).

As an example of the balancing of social values, Hafner and Berman discuss cases about hunting wild animals. In one case, the plaintiff was a fisherman closing his large net, whereupon the defendant entered through the remaining opening and caught the fish inside (Young v Hitchens 1844). Here there was a conflict between the competing social values of the pursuit of livelihood through productive work and economic competition. By deciding for the plaintiff or the defendant, a court can achieve the promotion of one value, but at the price of demoting the other. Here the court found for the defendant, but the judges' opinions show the careful balancing in the background (see Berman and Hafner 1993; Hafner and Berman 2002 for more). This case and the other wild animal cases have been extensively studied in Artificial Intelligence and Law.

Continuing from the work by Hafner and Berman, Bench-Capon developed his value-based argumentation frameworks (2003). He emphasises that it is not just logical soundness that settles arguments in courts. In fact, he writes that 'arguments [of both sides] may all be sound. But their arguments will not have equal value for the judge charged with deciding the case: the case will be decided by the judge preferring one argument over the other.' Here Bench-Capon explicitly connects to the legacy of Perelman, who in the treatise *The New Rhetoric* cowritten with Olbrechts-Tyteca noted that

> If men oppose each other concerning a decision to be taken, it is not because they commit some error of logic or calculation. They discuss apropos the applicable rule, the ends to be considered, the meaning to be given to values, the interpretation and characterisation of facts. (Perelman and Olbrechts-Tyteca 1969, as quoted by Bench-Capon 2003)

In his formal account of the role of values in argumentation, Bench-Capon starts from Dung's abstract argumentation frameworks (1995). An abstract argumentation framework is a directed graph, the nodes of which represent unstructured abstract arguments, while the edges express the attack relation between arguments. Bench-Capon's value-based argumentation frameworks consist of an abstract argumentation framework where each argument has an associated abstract value. The intended meaning is that when an argument is accepted, the value associated with the

argument is promoted. Bench-Capon continues to define audience-specific argumentation frameworks that consist of a value-based argumentation framework with a strict partial order on the values, representing an audience's preference relation on the values. Attack in the underlying abstract argumentation framework is connected to the values and their preferences by stipulating that an argument A defeats an argument B (for an audience a) if A attacks B and the value associated with B is not preferred to the value associated with A for audience a. The notion of defeat is then used in adaptations of Dung's abstract argumentation semantics. Bench-Capon continued his work on value-guided argumentation in a series of papers, often using Hafner and Berman's wild animal cases as background (Atkinson and Bench-Capon 2006; Bench-Capon et al. 2005; Bench-Capon and Sartor 2003).

Both these lines of research continue from work on case-based reasoning in the law (Aleven and Ashley 1995; Ashley 1990; Rissland 1983; Rissland and Ashley 1987, 2002). In this work, legal decision making is studied as guided by the principle of *stare decisis*, where decisions in new cases follow past cases. This work models how past cases are the source of hypothetical arguments that can guide decision making in a new case. A case can for instance be modeled as a set of factors, representing generalized facts for or against a decision. A past case is more on point than another when it shares more factors with the current case. In this way, the shared factors represent an analogy relation between cases, (partially) ordered by on-pointness. Cases can be distinguished from one another by noting factors that are not shared. Factors can range over dimensions, indicating a kind of strength of the factor.

3 Formalism

The following formal perspective has been developed in recent research on evidential argumentation, in order to bridge between qualitative or quantitative modeling primitives, in particular arguments, scenarios and probabilities (Verheij 2014, 2016c; Verheij et al. 2016), building on (Verheij 2010, 2012). In subsequent sections, we show how the formalism also can be put to work for ethical decision making and its context-dependence, value-dependence and rule-dependence.

3.1 General idea

The formalism models arguments that can be presumptive (also called ampliative), in the sense of logically going beyond their premises. Against the background of classical logic, an argument from premises P to conclusions Q goes beyond its premises when Q is not logically implied by P. Many arguments used in practice are presumptive. For instance, the prosecution may argue that a suspect was at the crime scene on the basis of a witness testimony. The fact that the witness has testified as such does not logically imply the fact that the suspect was at the crime scene. In particular, when the witness testimony is intentionally false, or is based on inaccurate observations or memories, the suspect may not have been at the crime scene at all. Denoting the witness testimony by P and the suspect being at the crime

scene as Q, the argument from P to Q is presumptive since P does not logically imply Q. For presumptive arguments, it is helpful to consider the *case made by the argument*, defined as the conjunction of the premises and conclusions of the argument (Verheij 2010, 2012). The case made by the argument from P to Q is $P \wedge Q$, using the conjunction of classical logic. An example of a non-presumptive argument goes from $P \wedge Q$ to Q. Here Q is logically implied by $P \wedge Q$. Presumptive arguments are often defeasible (Pollock 1987; Toulmin 1958), in the sense that extending the premises may lead to the retraction of conclusions.

Figure 1 shows two presumptive arguments from the same premises P: one supports the case $P \wedge Q$, the other the case $P \wedge \neg Q$. The $>$-sign indicates that one argument makes a stronger case than the other, resolving the conflict: the argument for the case $P \wedge Q$ is stronger than that for $P \wedge \neg Q$. The figure also shows two presumptions P and $\neg P$, treated as arguments from logically tautologous premises. Here the presumption $\neg P$ makes the stronger case when compared to the presumption P. Logically such presumptions can be treated as arguments from logical truth $\top$. The arguments make three cases: $\neg P$, $P \wedge Q$ and $P \wedge \neg Q$ (Fig. 2). The size of their areas suggest a preference relation.

The comparison of arguments and of cases are closely related in our approach, which can be illustrated as follows. The idea is that a case is preferred to another case if there is an argument with premises that supports the former case more strongly than the latter case. Hence, in the example in the figures, $\neg P$ is preferred to both $P \wedge Q$ and $P \wedge \neg Q$, and $P \wedge Q$ is preferred to $P \wedge \neg Q$. Conversely, given the cases and their preferences, we can compare arguments. The argument from P to Q is stronger than from P to $\neg Q$ when the best case that can be made from $P \wedge Q$ is preferred to the best case that can be made from $P \wedge \neg Q$.

3.2 Case models and arguments

We now formalize case models and how they can be used to interpret arguments. The formalism uses a classical logical language L generated from a set of propositional constants in a standard way. We write $\neg$ for negation, $\wedge$ for conjunction, $\vee$ for disjunction, $\leftrightarrow$ for equivalence, $\top$ for a tautology, and $\bot$ for a contradiction. The associated classical, deductive, monotonic consequence relation is denoted $\models$. We assume a finitely generated language.

First we define case models, formalizing the idea of cases and their preferences. The cases in a case model must be logically consistent, mutually incompatible and

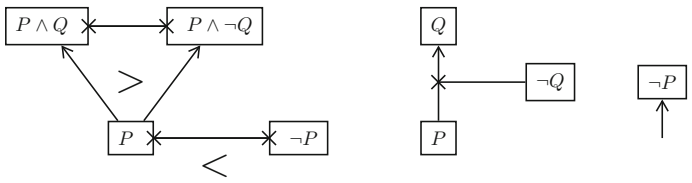

Fig. 1 Some arguments

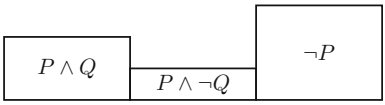

Fig. 2 Some cases

different; and the comparison relation must be total and transitive [hence is what is called a total preorder, commonly modeling preference relations (Roberts 1985)].

Definition 1 A *case model* is a pair $(C, \geq)$ with finite $C \subseteq L$, such that the following hold, for all φ, ψ and $\chi \in C$:

1. $\not\models \neg\varphi$;
2. If $\not\models \varphi \leftrightarrow \psi$, then $\models \neg(\varphi \wedge \psi)$;
3. If $\models \varphi \leftrightarrow \psi$, then $\varphi = \psi$;
4. $\varphi \geq \psi$ or $\psi \geq \varphi$;
5. If $\varphi \geq \psi$ and $\psi \geq \chi$, then $\varphi \geq \chi$.

The strict weak order $>$ standardly associated with a total preorder $\geq$ is defined as $\varphi > \psi$ if and only if it is not the case that $\psi \geq \varphi$ (for φ and $\psi \in C$). When $\varphi > \psi$, we say that φ is (strictly) preferred to ψ. The associated equivalence relation $\sim$ is defined as $\varphi \sim \psi$ if and only if $\varphi \geq \psi$ and $\psi \geq \varphi$.

Example 2 Figure 2 shows a case model with cases $\neg P$, $P \wedge Q$ and $P \wedge \neg Q$. $\neg P$ is (strictly) preferred to $P \wedge Q$, which in turn is preferred to $P \wedge \neg Q$.

Next we define arguments from premises $\varphi \in L$ to conclusions $\psi \in L$.

Definition 3 (*Arguments*) An *argument* is a pair (φ, ψ) with φ and $\psi \in L$. The sentence φ expresses the argument's premises, the sentence ψ its conclusions, and the sentence $\varphi \wedge \psi$ the *case made* by the argument. Generalizing, a sentence $\chi \in L$ is a *premise* of the argument when $\varphi \models \chi$, a *conclusion* when $\psi \models \chi$, and a *position* in the case made by the argument when $\varphi \wedge \psi \models \chi$. An argument (φ, ψ) is *(properly) presumptive* when $\varphi \not\models \psi$; otherwise *non-presumptive*. An argument (φ, ψ) is a *presumption* when $\models \varphi$, i.e., when its premises are logically tautologous.

Note our use of the plural for an argument's premises, conclusions and positions. This terminological convention allows us to speak of the premises $\mathtt{p}$ and $\neg\mathtt{q}$ and conclusions $\mathtt{r}$ and $\neg\mathtt{s}$ of the argument $(\mathtt{p} \wedge \neg\mathtt{q}, \mathtt{r} \wedge \neg\mathtt{s})$. Also the convention fits our non-syntactic definitions, where for instance an argument with premise χ also has logically equivalent sentences such as $\neg\neg\chi$ as a premise.

Coherent arguments are defined as arguments that make a case that is logically implied by a case in the case model.

Definition 4 (*Coherent arguments*) Let $(C, \geq)$ be a case model. Then we define, for all φ and $\psi \in L$:

$(C, \geq) \models (\varphi, \psi)$ if and only if $\exists\omega \in C$: $\omega \models \varphi \wedge \psi$.

We then say that the argument from φ to ψ is *coherent* with respect to the case model.

Conclusive arguments are defined as coherent arguments with the property that each case that implies the argument's premises also implies the argument's conclusions.

Definition 5 (*Conclusive arguments*) Let $(C, \geq)$ be a case model. Then we define, for all φ and $\psi \in L$:

$(C, \geq) \models \varphi \Rightarrow \psi$ if and only if $\exists \omega \in C$: $\omega \models \varphi \wedge \psi$ and $\forall \omega \in C$: if $\omega \models \varphi$, then $\omega \models \varphi \wedge \psi$.

We then say that the argument from φ to ψ is *conclusive* with respect to the case model.

Example 6 (continued from Example 2) In the case model of Fig. 2, the arguments from $\top$ to $\neg P$ and to P, and from P to Q and to $\neg Q$ are coherent and not conclusive in the sense of this definition. Denoting the case model as $(C, \geq)$, we have $(C, \geq) \models (\top, \neg P)$, $(C, \geq) \models (\top, P)$, $(C, \geq) \models (P, Q)$ and $(C, \geq) \models (P, \neg Q)$. The arguments from a case (in the case model) to itself, such as from $\neg P$ to $\neg P$, or from $P \wedge Q$ to $P \wedge Q$ are conclusive. The argument $(P \vee R, P)$ is also conclusive in this case model, since all $P \vee R$-cases are P-cases. Similarly, $(P \vee R, P \vee S)$ is conclusive.

The notion of presumptive validity considered here uses the idea that some arguments make a better case than other arguments from the same premises. More precisely, an argument is presumptively valid if there is a case implying the case made by the argument that is at least as preferred as all cases implying the premises.

Definition 7 (*Presumptively valid arguments*) Let $(C, \geq)$ be a case model. Then we define, for all φ and $\psi \in L$:

$(C, \geq) \models \varphi \rightsquigarrow \psi$ if and only if $\exists \omega \in C$:

1. $\omega \models \varphi \wedge \psi$; and
2. $\forall \omega' \in C$: if $\omega' \models \varphi$, then $\omega \geq \omega'$.

We then say that the argument from φ to ψ is *(presumptively) valid* with respect to the case model. A presumptively valid argument is *defeasible*, when it is not conclusive.

Circumstances χ are *defeating* when $(\varphi \wedge \chi, \psi)$ is not presumptively valid. Defeating circumstances are *rebutting* when $(\varphi \wedge \chi, \neg\psi)$ is presumptively valid; otherwise they are *undercutting*. Defeating circumstances are *excluding* when $(\varphi \wedge \chi, \psi)$ is not coherent.

Example 8 (continued from Examples 2 and 6) In the case model of Fig. 2, the arguments from $\top$ to $\neg P$, and from P to Q are presumptively valid in the sense of this definition. Denoting the case model as $(C, \geq)$, we have formally that $(C, \geq) \models \top \rightsquigarrow \neg P$ and $(C, \geq) \models P \rightsquigarrow Q$. The coherent arguments from $\top$ to P and from P to $\neg Q$ are not presumptively valid in this sense.

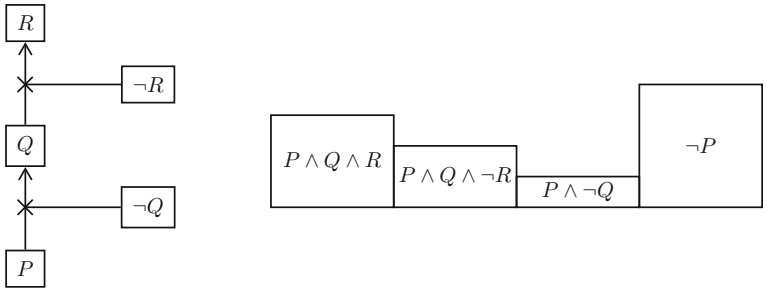

Fig. 3 An argument with two steps, each with exceptions

Example 9 Arguments typically consist of multiple steps. Figure 3 shows a two step argument on the left. The first step is from P to Q, the second from Q to R. Both steps have defeating circumstances: the first $\neg Q$, the second $\neg R$. On the right in the figure a case model is shown with the indicated arguments and defeating circumstances. The size of the areas of the boxes suggest the preference relation. From P, Q follows presumptively since $P \wedge Q \wedge R$ is a preferred case given P (in fact: the preferred case). From Q follows R. $\neg Q$ provides defeating circumstances for the former presumptive inference, since there is no preferred case of $P \wedge \neg Q$ in which R holds. (That preferred case is $P \wedge \neg Q$.) $\neg R$ gives defeating circumstances for the second presumptive inference. Formally, we have:

$$(C, \geq) \models P \rightsquigarrow Q$$
$$(C, \geq) \models Q \rightsquigarrow R$$
$$(C, \geq) \not\models P \wedge \neg Q \rightsquigarrow Q$$
$$(C, \geq) \not\models Q \wedge \neg R \rightsquigarrow R$$

Note that in the case model also the following hold:

$$(C, \geq) \models P \rightsquigarrow Q \wedge R$$
$$(C, \geq) \models Q \Rightarrow P$$
$$(C, \geq) \models R \Rightarrow Q$$
$$(C, \geq) \models R \Rightarrow P \wedge Q$$
$$(C, \geq) \models \top \rightsquigarrow \neg P$$

Figure 4 provides an ampliation version of the arguments, in which the cases made at each argumentative step are made explicit.

4 Dependence on contexts, values and rules

We now discuss the examples used in the introduction to illustrate the context-dependence, value-dependence and rule-dependence of ethical decision-making.

Example 10 (Context-dependence) Context-dependence was illustrated with the example that when there is a sudden risk of collision while driving on the highway, an abrupt stop can be a good idea, but not when there is someone close behind you.

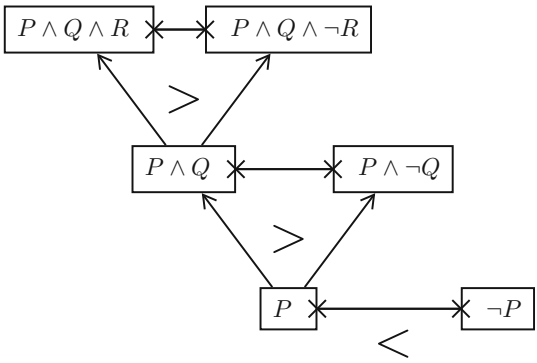

Fig. 4 An argument with two steps: ampliation version

Then it is better to slow down by careful braking. A case model $(C, \geq)$ for this example consists of three cases:

Case 1: CONTINUE-DRIVING $\wedge$ $\neg$ABRUPT-STOP $\wedge$ $\neg$CAREFUL-BREAKING
Case 2: $\neg$CONTINUE-DRIVING $\wedge$ ABRUPT-STOP $\wedge$ $\neg$CAREFUL-BREAKING $\wedge$ RISK-OF-COLLISION
Case 3: $\neg$CONTINUE-DRIVING $\wedge$ $\neg$ABRUPT-STOP $\wedge$ CAREFUL-BREAKING $\wedge$ RISK-OF-COLLISION $\wedge$ SOMEONE-CLOSE-BEHIND
Case 1 $>$ Case 2 $>$ Case 3

Case 1 is the normal situation of continuing to drive. It is the maximally preferred case, hence is the default situation:

$(C, \geq) \models \top \sim\!\!> \text{CONTINUE-DRIVING}$

It holds that RISK-OF-COLLISION presumptively implies ABRUPT-STOP, but not when also SOMEONE-CLOSE-BEHIND. Formally:

$(C, \geq) \models \text{RISK-OF-COLLISION} \sim\!\!> \text{ABRUPT-STOP}$
$(C, \geq) \not\models \text{RISK-OF-COLLISION} \wedge \text{SOMEONE-CLOSE-BEHIND} \sim\!\!> \text{ABRUPT-STOP}$
$(C, \geq) \models \text{RISK-OF-COLLISION} \wedge \text{SOMEONE-CLOSE-BEHIND} \sim\!\!> \neg\text{ABRUPT-STOP}$

Example 11 (Value-dependence) Value-dependence was illustrated with some drivers valuing speed more highly, and others safety. Assuming that maximizing the values of speed and safety are competing purposes to strive for (while driving), we can consider the following three cases in a case model.

Case 1: DRIVE $\wedge$ MAXIMIZE-SPEED $\wedge$ $\neg$MAXIMIZE-SAFETY
Case 2: DRIVE $\wedge$ $\neg$MAXIMIZE-SPEED $\wedge$ MAXIMIZE-SAFETY
Case 3: $\neg$DRIVE

The preference relation determines which choice is made. When the two cases are equally preferred, we have that both MAXIMIZE-SPEED and MAXIMIZE-SAFETY are presumptively valid conclusions. When Case 1 is preferred over the other, only

MAXIMIZE-SPEED presumptively follows; when Case 2 is preferred, only MAXIMIZE-SAFETY. Formally:

When Case 1 $\sim$ Case 2:
$(C, \geq) \models$ DRIVE $\rightsquigarrow$ MAXIMIZE-SPEED
$(C, \geq) \models$ DRIVE $\rightsquigarrow$ MAXIMIZE-SAFETY.
When Case 1 $>$ Case 2:
$(C, \geq) \models$ DRIVE $\rightsquigarrow$ MAXIMIZE-SPEED
$(C, \geq) \not\models$ DRIVE $\rightsquigarrow$ MAXIMIZE-SAFETY.
When Case 1 $<$ Case 2:
$(C, \geq) \not\models$ DRIVE $\rightsquigarrow$ MAXIMIZE-SPEED
$(C, \geq) \models$ DRIVE $\rightsquigarrow$ MAXIMIZE-SAFETY.

When Case 1 $\sim$ Case 2, it does not presumptively follow that MAXIMIZE-SPEED $\wedge$ MAXIMIZE-SAFETY showing that the (And)-rule does not hold for presumptive validity. When there is no preference for driving or not-driving, Case 3 is preferentially equivalent to both Case 1 and Case 2 (when they are equivalent) or to the preferred case (when one is preferred over the other).

Example 12 (Rule-dependence) Rule-dependence was illustrated with the reduced speed limit in residential areas. The following case model shows four different residential areas A, B, C and D and their speed limits.

Case 1: RESIDENTIAL $\wedge$ SPEED-LIMIT-30 $\wedge$ AREA-A
Case 2: RESIDENTIAL $\wedge$ SPEED-LIMIT-30 $\wedge$ AREA-B
Case 3: SPEED-LIMIT-30 $\wedge$ AREA-C
Case 4: SPEED-LIMIT-50 $\wedge$ AREA-D
Case 1 $\sim$ Case 2 $<$ Case 3 $\sim$ Case 4
Background theory: $\neg$(AREA-A $\wedge$ AREA-B) $\wedge$ $\neg$(AREA-A $\wedge$ AREA-C))
$\wedge$ $\neg$(AREA-A $\wedge$ AREA-D)) $\wedge$ $\neg$(AREA-*B* $\wedge$ AREA-C)
$\wedge$ $\neg$(AREA-B $\wedge$ AREA$-$D) $\wedge$ $\neg$(AREA-C $\wedge$ AREA-D)
$\wedge$ $\neg$(SPEED-LIMIT-30 $\wedge$ SPEED-LIMIT-50)

The preference relation is meant to suggest that the residential areas A and B are exceptional. A separate background theory sentence is specified that holds in all four cases. It expresses that the four residential areas are different and that there is only one speed limit. Here SPEED-LIMIT-30 follows presumptively (even conclusively) from AREA-A and from AREA-B. We find that in this case model the rule holds that in residential areas the speed limit is 30 km/h. The rule is both presumptively and conclusive valid:

$(C, \geq) \models$ RESIDENTIAL $\rightsquigarrow$ SPEED-LIMIT-30
$(C, \geq) \models$ RESIDENTIAL $\Rightarrow$ SPEED-LIMIT-30

The reversed rule with antecedent and consequent switched is not presumptively valid (hence also not conclusively):

$(C, \geq) \not\models$ SPEED-LIMIT-30 $\rightsquigarrow$ RESIDENTIAL

5 The development of the relevance of cases

A specific theme addressed by Hafner and Berman is that the relevance of a case as an authoritative source to base new decisions on can evolve over time (Berman and Hafner 1995; Hafner and Berman 2002). The precedential value is not cast in stone, but develops over time. As their main example, they discuss a series of New York tort cases about car accidents. The issue was whether a driver should repair a passenger's damages. The series of cases are about what should be done when different jurisdictions are relevant, each with a different authoritative solution. For instance, when the driver and passenger are from New York, where the trip starts, and the accident happens in Ontario, Canada, should then the Ontario rule be followed—barring a law suit in such a case—or the New York rule where negligent driving could imply recovery of damages? Hafner and Berman discuss a series of cases that show the tension between a territory perspective, where the location of the accident (the *situs*) is leading, and a forum perspective, where the place of litigation determines the applicable law. Gradually, the cases shift from a strict territorial rule to a center-of-gravity rule, where the circumstances are weighed.

Smith v Clute 277 N.Y. 407, 14 N.E.2d 455 (1938) *The claim was in tort law (driver negligence). The territorial rule applies.* The case concerned New Yorkers traveling in Montana. Following the territorial principle, Montana law was found applicable. Still the passenger won since he did not count as a guest passenger.

Kerfoot v Kelley 294 N.Y. 288, 62 N.E.2d 74 (1945) *The claim was in tort law (driver negligence). The territorial rule applies.* A case of traveling New Yorkers, now in Virginia. The passenger died in the accident. Again territory is determining, and Virginia law was found applicable. The driver won since the standard was gross negligence.

Auten v Auten 308 N.Y. 155, 124 N.E.2d 99 (1954) *The claim was in contract law (enforce a child support agreement). The center-of-gravity rule applies.* The execution of the contract in New York was fortuitous as the important contacts were mostly in England. The center of gravity approach was used to find England law applicable, but this is contract law, not tort law, where the territorial principle still reigns as we see in the next case.

Kaufman v American Youth Hostels 5 N.Y.2d 1016 (1959) *The claim was in tort law (travel guide negligence). The territorial rule applies.* A New York plaintiff, a New York defendant. A mountain climber died in an accident in Oregon. In Oregon, charities were immunized from tort liability because of wrongful death. Using the territory rule, the Oregon rule was found applicable.

Haag v Barnes 9 N.Y.2d 554, 175 N.E.2d 441, 216 N.Y.S. 2d 65 (1961) *The claim was in contract law (reopen a child support agreement). The center-of-gravity rule applies.* Meanwhile, in contract law, the center-of-gravity principle used in Auten is reinforced. A New York plaintiff, an Illinois defendant. Illinois law was found applicable since the agreement described the parties as being 'of Chicago, Illinois', the child was born in Illinois and the payments were made from Illinois. Center of gravity is Illinois, not New York.

Kilberg v Northeast Airlines 9 N.Y.2d 34, 172 N.E.2d 526, 211 N.Y.S.2d 133 (1961) *The claim was in tort law (common carrier negligence). The territorial rule is overridden for reasons of public policy.* The first crack in the territorial approach for tort law. Kilberg flew from New York, where the ticket was bought. The plane crashed in Nantucket, Massachusetts. The territorial rule would say Massachusetts law applies, but an exception is made. New York law was found applicable, by which the damages recoverable were unlimited, as opposed to the $15000 Massachusetts limit.

Babcock v Jackson 12 N.Y.2d 473, 191 N.E.2d 279, 473 N.Y.S.2d 279 (1963) *The claim was in tort law (driver negligence). The center-of-gravity rule applies.* A landmark case overriding previous cases, by which the center-of-gravity approach is established for tort law. Two New Yorkers drive to Ontario, Canada, where they have an accident, injuring Babcock. New York law only requires that negligent driving is shown, a rule that is found applicable using the center-of-gravity perspective (also referred to as 'grouping of contacts').[3]

In this series of cases, two factors stand out when analyzing the development of their precedential relevance: the kind of case and the jurisdiction choice rule. Is the case a tort case (TORT) or a contract case (CONTRACT)? And: Was the territorial rule applied (TERRITORY), was an exception to its validity being made (EXCEPTION), or was the center-of-gravity rule applied (GRAVITY)? Figure 5 summarizes these factors for the cases listed. The cases are identified by factors for the plaintiff's name and the year of the decision.

Now consider the following formal case model, consisting of 7 cases:

SMITH $\wedge$ 1938 $\wedge$ TORT $\wedge$ TERRITORY
KERFOOT $\wedge$ 1945 $\wedge$ TORT $\wedge$ TERRITORY
AUTEN $\wedge$ 1954 $\wedge$ CONTRACT $\wedge$ GRAVITY
KAUFMAN $\wedge$ 1959 $\wedge$ TORT $\wedge$ TERRITORY
HAAG $\wedge$ 1961 $\wedge$ CONTRACT $\wedge$ GRAVITY
KILBERG $\wedge$ 1961 $\wedge$ TORT $\wedge$ EXCEPTION
BABCOCK $\wedge$ 1963 $\wedge$ TORT $\wedge$ GRAVITY

We assume a background theory holding in all cases in which the plaintiff names exclude each other pairwise ($\neg$(SMITH $\wedge$ KERFOOT), etc.), and similarly for the decision years ($\neg$(1938 $\wedge$ 1945), etc.), the kinds of cases ($\neg$(TORT $\wedge$ CONTRACT)) and the choice rules ($\neg$(TERRITORY $\wedge$ EXCEPTION), etc.). As preference ordering, we take all cases to be preferentially equivalent, except for the landmark Babcock case, which is preferred over the other cases.

[3] As the court says in this landmark case: 'Comparison of the relative 'contacts' and 'interests' of New York and Ontario in this litigation, vis–à–vis the issue here presented, makes it clear that the concern of New York is unquestionably the greater ...The present action involves injuries sustained by a New York guest as the result of the negligence of a New York host in the operation of an automobile, garaged, licensed and undoubtedly insured in New York, in the course of a week-end journey which began and was to end there. In sharp contrast, Ontario's sole relationship with the occurrence is the purely adventitious circumstance that the accident occurred there.' [Babcock, p. 458; as quoted by Hafner and Berman (2002)].

SMITH 1938 TORT TERRITORY	KERFOOT 1945 TORT TERRITORY	AUTEN 1954 CONTRACT GRAVITY	KAUFMAN 1959 TORT TERRITORY
HAAG 1961 CONTRACT GRAVITY	KILBERG 1961 TORT EXCEPTION	BABCOCK 1963 TORT GRAVITY	

Fig. 5 The evolving relevance of cases

We can now analyze the development of the jurisdiction choice rule by restricting the case model to the cases up and until a particular year. For instance, we write $C(1954)$ for the set consisting of the three cases Smith, Kerfoot and Auten dating from 1954 or before. Here is what we find.

In 1938, when only the Smith case is relevant, there is only one possibility. The territory rule holds in tort law cases. Formally, the case model is restricted to the singleton set $C(1938)$, and TERRITORY follows presumptively and conclusively from TORT:

$(C(1938), \geq) \models \text{TORT} \rightsquigarrow \text{TERRITORY}$
$(C(1938), \geq) \models \text{TORT} \Rightarrow \text{TERRITORY}$

Nothing changes up to and including the Kaufman case. The Auten case makes a change for contract cases, but not for tort cases. Also in 1959 TERRITORY follows presumptively and conclusively from TORT, as we see in the case model restricted to the set $C(1959)$ consisting of the four cases Smith, Kerfoot, Auten and Kaufman. The Haag case reenforces the center-of-gravity rule for contract cases, but the first sign of real change for tort cases occurs in 1961 with the Kilberg case, where a new possibility is enforced: it is possible that there is an exception to the validity of the territory rule. TERRITORY still follows presumptively from TORT, but no longer conclusively, as there is an alternative presumptive consequence EXCEPTION.

$(C(1961), \geq) \models \text{TORT} \rightsquigarrow \text{TERRITORY}$
$(C(1961), \geq) \models \text{TORT} \rightsquigarrow \text{EXCEPTION}$
$(C(1961), \geq) \not\models \text{TORT} \Rightarrow \text{TERRITORY}$

And then the landmark decision Babcock comes about in 1963. The precedential value of the territory rule tort cases (Smith, Kerfoot and Kaufman) is now limited by the preferred Babcock case. TERRITORY no longer follows presumptively from TORT, nor does EXCEPTION. From now on, it is only GRAVITY that follows presumptively from TORT:

$(C(1963), \geq) \not\models \text{TORT} \rightsquigarrow \text{TERRITORY}$
$(C(1963), \geq) \not\models \text{TORT} \rightsquigarrow \text{EXCEPTION}$
$(C(1963), \geq) \models \text{TORT} \rightsquigarrow \text{GRAVITY}$

In the present model, GRAVITY does not follow conclusively from TORT. This is as it should be since GRAVITY does not even follow presumptively in the older cases, for instance not in Smith:

$(C(1963), \geq) \not\models \text{TORT} \Rightarrow \text{GRAVITY}$
$(C(1963), \geq) \not\models \text{TORT} \wedge \text{SMITH} \rightsquigarrow \text{GRAVITY}$

As Hafner and Berman discuss in their work, the contract cases have an indirect influence on the development of jurisdiction choice in tort cases. They do not give rise to new possibilities in tort cases, but they do give new general possibilities. For instance, after the 1954 Auten case, we find that the gravity rule now presumptively follows in general—it has become a hypothetical possibility—but not in a tort case:

$(C(1954), \geq) \models \top \rightsquigarrow \text{GRAVITY}$
$(C(1954), \geq) \not\models \text{TORT} \rightsquigarrow \text{GRAVITY}$

This illustrates the hypothetical argumentation associated with case-based reasoning in the law, where different kinds of cases can give rise to new hypothetical possibilities to be further argued about (cf. the idea of cases as sources of hypotheticals, developed in case-based reasoning research in AI and Law. See Rissland (2013) for a historical account of the development of this idea, inspired by Lakatos' discussion of examples in mathematics).

Our formal treatment can be compared to the five temporal patterns signifying a weakening of precedent, as distinguished by Hafner and Berman (2002, p. 40). We go from slight weakening to strong weakening:

1. *A general shift in the relative priority of competing purposes.* Auten and Haag show that the territory rule loses the solid ground it had before given Smith and Kerfoot, although these cases do not yet have precedential value in tort cases. Formally, we saw that in 1954 GRAVITY became a presumptive conclusion in general—one of the options next to TERRITORY—whereas before it was not even a coherent position. However, in the more specific setting of TORT-cases GRAVITY was not yet a presumptive conclusion.
2. *A shift in the relative priority of competing purposes by finding exceptions.* Here the Kilberg case is an example, and indeed since 1961, TERRITORY is no longer a conclusive consequence of TORT, it is only one presumptive consequence, next to EXCEPTION.
3. *The ratio decidendi of an older case is overruled, although it is significantly different.* Here the Kaufman example is used as an illustration. Babcock overrules Kaufman in this sense, as Kaufman is not a passenger case. Our formal case model does not distinguish tort cases from passenger cases, so this specific kind of weakening is not here visible. An adaptation of the model can make the distinction.
4. *A case is implicitly overruled.* One can say that both Babcock and Kilberg implicitly overruled Kerfoot: had Kerfoot been decided after Babcock or Kilberg it might have been decided differently. Formally, in TORT cases after 1963, TERRITORY is no longer a presumptive conclusion and GRAVITY is.

5. *A case is explictly overruled.* Then it is explicitly stated that an earlier case's precedential value is overruled by the new case. As Hafner and Berman say, this occurs rarely. None of the discussed cases falls in this class. The difference between implicit and explicit overruling is not modeled in our formalism since we abstract from explicit references to cases.

6 A hierarchy of values

Hafner and Berman use hierarchical relationships between the values promoted and demoted by a choice of action. For the choice of law cases discussed in the previous section, they consider a hierarchy consisting of eight elements. Figure 6 is an adapted version of their Figure 7a (Hafner and Berman 2002, p. 41). In the figure, they distinguish two competing choices of action: using the territorial rule (PA), and using the center-of-gravity rule (PB). The former promotes predictability (PP) and avoids forum shopping (AS), which in turn promotes efficiency (PE). The latter avoids fortuitousness (AF), protects the interests of states (PI) and in particular their regulation interests (PR).

We develop a formal case model that represents the hierarchy. The first step is to distinguish two kinds of cases: cases in which the territorial rule is used (PA) and cases in which the center-of-gravity rule is used (PB). Hafner and Berman model these as being separate, so PA and PB exclude each other. Formally, PA cases are ¬PB cases, and PB cases have ¬PA.

The hierarchy shows, how in PA cases, several purposes are achieved: predictability is promoted (PP), fortuitousness is avoided (AS) and efficiency is promoted (PE). So in a PA case the conjunction of PP, AS and PE follows. Hafner and Berman seem to consider these as conclusive consequences: The territorial rule always promotes predictability, i.e, there are no PA cases in which PP does not follow. Similarly for PB cases, where AF, PI and PR follow conclusively.

As a result, we have two kinds of cases:

$$\mathrm{PA} \wedge \neg \mathrm{PB} \wedge \mathrm{PP} \wedge \mathrm{AS} \wedge \mathrm{PE}$$
$$\neg \mathrm{PA} \wedge \mathrm{PB} \wedge \mathrm{AF} \wedge \mathrm{PI} \wedge \mathrm{PR}$$

In a case model with these two cases, and a trivial preference relation in which the cases are equivalent, we find that the arrows in the figure are all conclusively valid

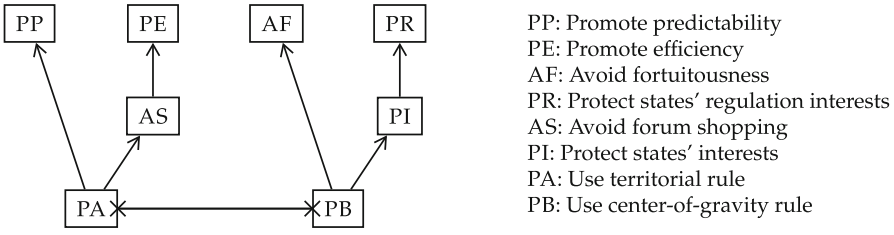

Fig. 6 A hierarchy of values promoted and demoted

inferences. For instance, PA conclusively implies PP and AS, and AS conclusively implies PE. Also PA conclusively implies ¬PB, and PB ¬PA, representing that they are exclusionary types. But in such a simple case model the hierarchy itself is not preserved as there are many more conclusively valid inferences. For instance, PP conclusively implies PA, AS, PE and −PB.

For representing the hierarchical structure, more cases are needed. For instance, to represent that PA cases are AS cases, but not vice versa, we need a case in which ¬PA. These cases will also be PE (since AS is a kind of PE) and ¬PB (since we need to distinguish from the other cases and the use of the gravity rule does not promote predictability):

¬PA ∧ ¬PB ∧ AS ∧ PE

This case allows for situations in which efficiency is promoted by avoiding forum shopping other than by using the territorial rule. Adding this case to the case model (and keeping the preference relation trivial with all cases equivalent), we find that PA still conclusively implies AS, whereas AS now implies PA only presumptively.

In order to also allow for PE cases that are not AS cases, i.e., cases with other kinds of efficiency promotion than avoiding forum shopping, we add a case in which ¬AS.

¬PA ∧ ¬PB ∧ ¬AS ∧ PE

Now AS still conclusively implies PE, but PE implies AS only presumptively.

Also there can be situations where predictability is promoted other than by using the territorial rule, so we need cases with PP and ¬PA. Such cases will also be ¬PB. A first idea is to include this case:

¬PA ∧ ¬PB ∧ PP

But this case is not distinguishable from the previous two cases (cf. the requirement that cases are mutually exclusive in a case model). One way of repairing this is by splitting the previous two cases into PP and ¬PP versions:

¬PA ∧ ¬PB ∧ AS ∧ PE ∧ PP
¬PA ∧ ¬PB ∧ AS ∧ PE ∧ ¬PP
¬PA ∧ ¬PB ∧ ¬AS ∧ PE ∧ PP
¬PA ∧ ¬PB ∧ ¬AS ∧ PE ∧ ¬PP

Doing the same for the PB side of the hierarchy (using the symmetry of the diagram), we get the following 10 cases:

PA ∧ ¬PB ∧ PP ∧ AS ∧ PE
¬PA ∧ PB ∧ AF ∧ PI ∧ PR
¬PA ∧ ¬PB ∧ AS ∧ PE ∧ PP
¬PA ∧ ¬PB ∧ AS ∧ PE ∧ ¬PP
¬PA ∧ ¬PB ∧ ¬AS ∧ PE ∧ PP
¬PA ∧ ¬PB ∧ ¬AS ∧ PE ∧ ¬PP

$\neg$PA $\wedge$ $\neg$PB $\wedge$ PI $\wedge$ PR $\wedge$ AF
$\neg$PA $\wedge$ $\neg$PB $\wedge$ PI $\wedge$ PR $\wedge$ $\neg$AF
$\neg$PA $\wedge$ $\neg$PB $\wedge$ $\neg$PI $\wedge$ PR $\wedge$ AF
$\neg$PA $\wedge$ $\neg$PB $\wedge$ $\neg$PR $\wedge$ PR $\wedge$ $\neg$AF

In the case model $(C, \geq)$ consisting of these 10 cases, all preferentially equivalent, we formally have the following, representing the model in the figure:

$(C, \geq) \models$ PA $\Rightarrow$ $\neg$PB $(C, \geq) \models$ PB $\Rightarrow$ $\neg$PA
$(C, \geq) \models$ PA $\Rightarrow$ PP $(C, \geq) \models$ PP $\rightsquigarrow$ PA
$(C, \geq) \models$ PA $\Rightarrow$ AS $(C, \geq) \models$ AS $\rightsquigarrow$ PA
$(C, \geq) \models$ AS $\Rightarrow$ PE $(C, \geq) \models$ PE $\rightsquigarrow$ AS
$(C, \geq) \models$ PB $\Rightarrow$ AF $(C, \geq) \models$ AF $\rightsquigarrow$ PB
$(C, \geq) \models$ PB $\Rightarrow$ PI $(C, \geq) \models$ PI $\rightsquigarrow$ PB
$(C, \geq) \models$ PI $\Rightarrow$ PR $(C, \geq) \models$ PR $\rightsquigarrow$ PI

7 Discussion

We have studied decision making and its dependence on contexts, values and rules. Contexts are present in our use of formalized cases, that can be considered as representing the relevant properties of a situation, possible or real. The values appear in the preference ordering on the cases in case models. The preferences help to make a choice of maximal value. The role of rules comes about when we consider how case models give rise to notions of presumptively and conclusively valid arguments with a conditional form.

We studied the cases illustrating the evolution of the relevance of cases studied by Hafner and Berman (Berman and Hafner 1995; Hafner and Berman 2002) in terms of our case models. We showed how the conclusively and presumptively valid choices developed over time and how they differ from context to context. We also showed how hierarchical relations between values promoted and demoted can be represented in a case model. In this way, we have provided a formal version of these cases, illustrating some of Hafner and Berman's central concerns, in particular context-dependence, value-dependence and rule-dependence.

Bench-Capon built his value-based argumentation frameworks on top of Dung's abstract argumentation, a natural choice by the innovative technical possibilities allowed by that formalism. Our approach is not based on abstract argumentation, but has been developed in a way to stay close to classical logic and standard probability theory (see Verheij 2012, 2014, 2016b, c). Bench-Capon modeled the promotion and demotion of values as an argument selection mechanism. In our model, the promotion and demotion of values appears in the arguments that are conclusively and presumptively valid given the premises. For instance, we saw how the choice for the territory rule (PA) promoted predictability (PP) and efficiency (PE). Formally, PA$\rightsquigarrow$PP and PA$\rightsquigarrow$PE were presumptively valid, and in fact conclusive (PA$\Rightarrow$PP and PA$\Rightarrow$PE) in the model developed.

Here we have not addressed reasoning about values, as we did in (Verheij 2013). There we built on a different kind of argumentation formalism (DefLog), a model extending Dung's abstract argumentation with support and with support/attack about support/attack by the use of nested conditionals. Here we have not included such reasoning in our discussions. It can be noted that nested conditionals such as $P \rightarrow (Q \rightarrow R)$ play a role in reasoning that is in relevant ways similar to the conditional with a composite antecedent $P \wedge Q \rightarrow R$. Concretely, for the nested conditional and for the conditional-with-composite-antecedent, one expects that when both P and Q hold, R follows. The conditional-with-composite-antecedent has been studied in the present paper, in its presumptive and conclusive forms $P \wedge Q \rightsquigarrow R$ and $P \wedge Q \Rightarrow R$. One idea would be to define $P \rightsquigarrow (Q \rightsquigarrow R)$ and $P \Rightarrow (Q \Rightarrow R)$ as these conditionals-with-composite-antecedent. In collaboration with Modgil, Bench-Capon has developed his value-based argumentation frameworks to the modeling of arguments about value preferences (Bench-Capon and Modgil 2009; Modgil and Bench-Capon 2011). In contrast with these models, the present stays close to logic and probability theory, whereas they work with adaptations of abstract argumentation.

Another kind of model has been developed by Atkinson and Bench-Capon who focused on practical reasoning about which actions to choose (Atkinson and Bench-Capon 2006, 2007), where they use Belief–Desire–Intention (BDI) modeling, Action-Based Alternating Transition Systems (AATS) and argumentation schemes . These approaches are very relevant for the present work, now that the kind of decision making studied here has close similarities to practical reasoning. However, intentional aspects (associated with BDI modeling), coordination between agents (as studied in AATS modeling) and dialogical themes (as they naturally arise when studying argument schemes and their critical questions) are beyond the scope of the present abstract model.

By the use of case models and the study of evolving precedential value, the present work has connections to case-based reasoning in the law more generally. For instance, there are clear connections to case-based reasoning in the law (Aleven and Ashley 1995; Ashley 1990; Rissland 1983; Rissland and Ashley 1987, 2002). The elementary propositions of the logical language used to express the cases in our case models are closely related to factors, although the latter are pro-plaintiff or pro-defendant, and ours are not. Whether a proposition is pro-plaintiff or pro-defendant would have to be determined on the basis of other information in the case model. For instance, if a factor F is pro-plaintiff P, this can be thought of as the conditional $F \Rightarrow P$ being valid in the case model. Or, allowing for a factor being hypothetically for a side in the debate, $F \rightsquigarrow P$ could be valid. Our approach does not distinguish the dimensionality that can come with factors, although dimensions add significantly to the expressiveness and relevance of a set of modeling tools. Since our model is connected to the bridging of qualitative and quantitative modeling primitives, it may be interesting to include dimensions in the model. A key difference between the model in the present paper and the listed work on case-based reasoning in the law is that we stay close to logic and probability theory, and develop a theory of conclusive and presumptive validity.

8 Concluding remarks

The paper started with the ethical dimension of AI, and discussed how advances in technology necessitate that systems develop to full ethical systems, in the sense that they can make decisions while taking the relevant context, human values and normative rules into account.

We discussed AI&Law research involving value-guided argumentation and used that as an inspiration for discussing dependence on contexts, values and rules. Examples by Hafner and Berman were analyzed. We studied the development of the relevance of cases and the hierarchy of values. In this way we showed how a formalism developed for bridging qualitative and quantitative primitives in evidential reasoning can be applied to value-guided argumentation grounded in cases.

The results are relevant for ethical system design, as one way of looking at ethical system design is as technology that is better suited for who we are as humans. A simple example could be a smartphone that does not make sounds during the times that we are supposed to be sleeping, or better yet: that does not give immediate access to email and facebook during those times. Such interruptions can be fine, and can under circumstances even be rational, but most often it is best to sleep at night. Autonomous driving requires ethical decisions of significantly greater complexity. Always ethical systems should be aware of their relevant context, have embedded values, and use the rules that apply in order to do what is right. Ethical system design is the way of the future, and here some suggestions have been made for their formal foundations.

References

Aleven V, Ashley KD (1995) Doing things with factors. In: Proceedings of the fifth international conference on artificial intelligence and law (ICAIL'1995). ACM Press, New York, pp 31–41

Ashley KD (1990) Modeling legal arguments: reasoning with cases and hypotheticals. The MIT Press, Cambridge

Atkinson K, Bench-Capon TJM (2006) Legal case-based reasoning as practical reasoning. Artif Intell Law 13:93–131

Atkinson K, Bench-Capon TJM (2007) Practical reasoning as presumptive argumentation using action based alternating transition systems. Artif Intell 171:855–874

Bench-Capon TJM (2003) Persuasion in practical argument using value-based argumentation frameworks. J Logic Comput 13(3):429–448

Bench-Capon TJM, Atkinson K, Chorley A (2005) Persuasion and value in legal argument. J Logic Comput 15(6):1075–1097

Bench-Capon TJM, Modgil S (2009) Case law in extended argumentation frameworks. In: Proceedings of the 12th international conference on artificial intelligence and law (ICAIL 2009). ACM Press, New York, pp 118–127

Bench-Capon TJM, Sartor G (2003) A model of legal reasoning with cases incorporating theories and values. Artif Intell 150(1):97–143

Berman DH, Hafner CL (1991) Incorporating procedural context into a model of case-based legal reasoning. In: Proceedings of the third international conference on artificial intelligence and law. ACM Press, New York, pp 12–20
Berman DH, Hafner CL (1993) Representing teleological structure in case based reasoning: the missing link. In: Proceedings of the fourth international conference on artificial intelligence and law. ACM Press, New York, pp 50–59
Berman DH, Hafner CL (1995) Understanding precedents in a temporal context of evolving legal doctrine. In: Proceedings of the fifth international conference on artificial intelligence and law. ACM Press, New York, pp 42–51
Briggs R (2014) Normative theories of rational choice: expected utility. In: Zalta EN (ed) The Stanford encyclopedia of philosophy. Stanford University, Stanford
Broersen J (2014) Responsible intelligent systems. Künstl Intell 28:209–214
Dawid A (1987) The difficulty about conjunction. J R Stat Soc Ser D (Stat) 36(2/3):91–92
Dung PM (1995) On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif Intell 77:321–357
Hafner CL, Berman DH (2002) The role of context in case-based legal reasoning: teleological, temporal, and procedural. Artif Intell Law 10(1–3):19–64
Modgil S, Bench-Capon TJM (2011) Metalevel argumentation. J Logic Comput 21(6):959–1003
Moor JH (2006) The nature, importance, and difficulty of machine ethics. IEEE Intell Syst 21:18–21
Perelman C, Olbrechts-Tyteca L (1958/1969) The new rhetoric: a treatise on argumentation [English translation of La nouvelle rhétorique: Traité de l'argumentation]
Pollock JL (1987) Defeasible reasoning. Cogn Sci 11(4):481–518
Rissland EL (1983) Examples in legal reasoning: legal hypotheticals. In: Proceedings of the 8th international joint conference on artificial intelligence (IJCAI'1983), pp 90–93
Rissland EL (2013) From UUM and CEG to CBR and ICAIL: a journey in AI and law. In: From knowledge representation to argumentation in AI, law and policy making. A festschrift in honour of Trevor Bench-Capon on the occasion of his 60th birthday. College Publications, London, pp 191–212
Rissland EL, Ashley KD (1987) A case-based system for trade secrets law. In: Proceedings of the first international conference on artificial intelligence and law. ACM Press, New York, pp 60–66
Rissland EL, Ashley KD (2002) A note on dimensions and factors. Artif Intell Law 10:65–77
Roberts FS (1985) Measurement theory with applications to decisionmaking, utility, and the social sciences. Cambridge University Press, Cambridge
Toulmin SE (1958) The uses of argument. Cambridge University Press, Cambridge
Verheij B (2010) Argumentation and rules with exceptions. In: Baroni B, Cerutti F, Giacomin M, Simari GR (eds) Computational models of argument: proceedings of COMMA 2010, Desenzano del Garda, Italy, Sept 8–10, 2010. IOS Press, Amsterdam, pp 455–462
Verheij B (2012) Jumping to conclusions: a logico-probabilistic foundation for defeasible rule-based arguments. In: Fariñas del Cerro L, Herzig A, Mengin J (eds) 13th European conference on logics in artificial intelligence (JELIA'2012). Toulouse, France. Proceedings (LNAI'7519). Springer, Berlin, pp 411–423
Verheij B (2013) Arguments about values. In: From knowledge representation to argumentation in AI, law and policy making. A festschrift in honour of Trevor Bench-Capon on the occasion of his 60th birthday. College Publications, London, pp 243–257
Verheij B (2014) To catch a thief with and without numbers: arguments, scenarios and probabilities in evidential reasoning. Law Probab Risk 13:307–325
Verheij B (2016) Arguments for ethical systems design. In: Bex FJ (ed) Legal knowledge and information systems: JURIX 2016. The twenty-ninth annual conference. IOS Press, Amsterdam
Verheij B (2016b) Correct grounded reasoning with presumptive arguments. In: Michael L, Kakas A (eds) 15th European conference on logics in artificial intelligence (JELIA'2016). Larnaca, Cyprus. Proceedings (LNAI'10021) Nov 9–11, 2016. Springer, Berlin
Verheij B (2016c) Formalizing correct evidential reasoning with arguments, scenarios and probabilities. In: Proceedings of the workshop on artificial intelligence for justice (AI4J 2016) at ECAI 2016, pp 87–95
Verheij B, Bex FJ, Timmer ST, Vlek CS, Meyer JJ, Renooij S, Prakken H (2016) Arguments, scenarios and probabilities: connections between three normative frameworks for evidential reasoning. Law Probab Risk 15:35–70